

# Bringing IATE into the Semantic Web Family

**Paula Diez-Ibarbia**  
Universidad Politécnica  
de Madrid  
paula.diez@upm.es

**Patricia Martín-Chozas**  
Universidad Politécnica  
de Madrid  
patricia.martin@upm.es

**Elena Montiel-Ponsoda**  
Universidad Politécnica  
de Madrid  
elena.montiel@upm.es

## Abstract

This paper is an extension of previous work by the authors and other researchers that studies the application of the OntoLex-lemon model for representing the InterActive Terminology for Europe (IATE) database in the Semantic Web. While traditional XML-based approaches have been effective for multilingual terminological work, the Semantic Web enables richer, more interoperable representations. The study evaluates the suitability of OntoLex-lemon for modeling IATE's complex structure and identifies limitations in existing vocabularies. To address these, this paper tries to identify other existing vocabularies and ontologies that could satisfy those limitations, which include term reliability, regional usage, lifecycle statuses, lookup forms, and concept cross-references. Still, some representation requirements are not covered by existing vocabularies and may need to be further discussed within the community.

## 1 Introduction

Traditional computational formats to structure terminological resources, such as those based on XML, have proven effective in supporting multilingual terminological work within commercial and industrial settings. These standards enable terminology teams to enforce consistent terminology and communication clarity while reducing time and cost.

Despite these strengths, representing terminologies using Semantic Web principles offers significant additional benefits. Unlike rigid XML hierarchies, Semantic Web standards such as RDF support more flexible, graph-based structures that facilitate incremental growth and easier integration. Furthermore, they enable interoperability across diverse datasets regardless of their origin, providing robust mechanisms for linking terminological data across languages and resources, enhancing reuse and connectivity in ways that traditional formats cannot easily achieve.

Among the Semantic Web models for representing terminologies and language resources, SKOS<sup>1</sup> and OntoLex-lemon<sup>2</sup> (hereforth, Ontolex) (McCrae et al., 2017) are the most widely adopted. The SKOS model has been successfully applied to large-scale thesauri such as EuroVoc (Díez et al., 2010), UNESCO Thesaurus (Pastor-Sánchez, 2016), AGROVOC (Caracciolo et al., 2013), TheSoz (Zapilko et al., 2013), and STW Thesaurus for Economics (Neubert, 2009).

In contrast to SKOS, Ontolex offers a wider linguistic framework as it provides a standardized model for representing lexical information. In fact, it has been widely applied to model lexical resources, such as the Apertium dictionaries (Gracia et al., 2018) and K Dictionaries (Bosque-Gil et al., 2016a). However, the potential of this model has also been thoroughly studied to represent terminological resources (Martín-Chozas et al., 2024).

In this line, the most representative and widely exploited terminological database in the European Union is the InterActive Terminology for Europe (IATE)<sup>3</sup>, which is the object of this study.

## 2 IATE in the Semantic Web

IATE, now IATE2 (Zorrilla-Agut and Fontenelle, 2019), is the official terminology database of the European Union. Its primary purpose is to promote clarity and accuracy in the drafting and translation of EU documents, and it is freely accessible to the public. Developed collaboratively by EU bodies and maintained by the Translation Centre for the Bodies of the EU, IATE contains millions of entries covering a wide range of domains, from law and finance to agriculture and science. The entries registered in this database can offer a wide range of information, not only terminological, but also

<sup>1</sup><http://www.w3.org/2004/02/skos/core#>

<sup>2</sup><https://www.w3.org/2016/05/ontolex/#core>

<sup>3</sup><https://iate.europa.eu/home>

lexicographical and conceptual.

The representation of this resource has already been addressed in the literature. First, Cimiano et al. (2015) proposed the representation of an IATE dump following the lemon vocabulary (foundation for the later Ontolex model) (McCrae et al., 2012) and complementary RDF properties from the TBX (Term Base eXchange) format, which was afterwards integrated in the TerminotecaRDF project (Bosque-Gil et al., 2016b), a platform to integrate semantically published terminological resources in Spanish. This first conversion of IATE was also the object of another work that aimed to enrich this resource with automatic translations (Arcan et al., 2018). This first attempt by Cimiano et al. (2015) to represent IATE in RDF was further extended by the creation of Terme-à-LLOD, a platform to convert and host terminological data based on TBX2RDF (di Buono et al., 2020).

Given this context, in this paper we explore the potential of Ontolex to meet the representation needs of IATE, identifying limitations of this standard and proposing complementary vocabularies to fill such void.

### 3 Representation requirements of IATE

As mentioned in the Introduction, the potential of Ontolex to represent terminological resources has previously been analysed by the authors in Martín-Chozas et al.(2024). This work revised a set of authoritative terminological resources, including IATE, and proposed an extension for the Ontolex model, Termlex<sup>4</sup>. Some of the requirements reported, that may require further discussion, are as follows:

- The definitions and notes of a term, which often include additional data, such as the author or the source. To accommodate these requirements, we proposed the classes termlex:Definition and termlex:Source.
- The reliability of a term, which indicates the accuracy or the level of confidence of a given term. This factor varies amongst resources. For instance, in IATE is represented with stars, from one (lower reliability) to four. In other resources, such as Termium, this is represented as an *acceptability rating*,

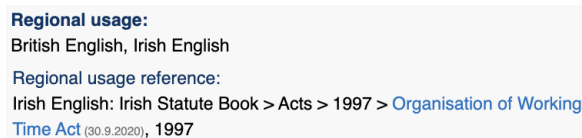
<sup>4</sup><https://termlex.oeg.fi.upm.es/myDocumentation3/index-en.html>

with values such as *correct*, *avoid* or *unofficial*. This indicator could be represented with `tbx:reliabilityCode`, but we believe that it should be standardised, as this scale varies amongst resources, and for this reason we proposed the `termlex:ReliabilityCode` class pointing at a fixed set of values.

Continuing with this work, in the following sections we propose newly identified elements that need further discussion.

#### 3.1 Regional Usage

A common element across IATE entries is the *regional use* indicator, as in Figure 1. This would not represent an issue itself, since Lexinfo<sup>5</sup> already provides support for this element with the properties `lexinfo:geographic` and `lexinfo:geographicVariant`. However, as shown in the figure, there are also references attached to this marker. Therefore, it may be required to reify this property to a class, so that the source information could be represented.



The image shows a screenshot of a web page with the following text:  
**Regional usage:**  
British English, Irish English  
**Regional usage reference:**  
Irish English: Irish Statute Book > Acts > 1997 > Organisation of Working Time Act (30.9.2020), 1997

Figure 1: Regional use for the entry *Pact for Skills*

#### 3.2 Lifecycle

Another feature present in some IATE entries is the *lifecycle* (Figure 2). This indicator can adopt four values: *historical* (no longer in use or in existence), *proposed* (but not yet adopted), or *abandoned* (proposed but ultimately not adopted), and it is used to mark the status of a terminological entry. A close property to model this is `lexinfo:normativeAuthorization`. However, its range acquires a fixed set of values (*admitted*, *deprecated*, *preferred*, etc.) that do not match the values of IATE. Moreover, this property is indicated to represent the status of a specific term, and the lifecycle indicator refers to the entire terminological entry. Therefore, this issue requires further discussion.

#### 3.3 Lookup Form

The *lookup form* is an interesting element that refers to any term or spelling variation that is

<sup>5</sup><https://lexinfo.net/>

★ IATE ID: 114385 Historical

**Domain:** national currency [ FINANCE > monetary economics > money market > money ]  
 monetary relations [ FINANCE > monetary relations ]

**Origins:**  
 Origin: (ES) Spain

**Cross-references:**  
 is related to: 861168 Kingdom of Spain

**Owner:** Consilium

Figure 2: Lifecycle for the entry *Peseta*

**Lookup form:** **E**

Term reference:

**Term type:** lookup form

**Reliability:** ★

**Owner:** Consilium

---

**Lookup form:** **ESP**

Term reference:

**Term type:** lookup form

**Reliability:** ★

**Owner:** Consilium

Figure 3: Lookup form for the entry *Reino de España*

searchable, but not displayed as a term, such as common spelling mistakes, alternative spellings, plural or inflected forms, etc. Figure 3 shows an example of two lookup forms for *Reino de España*. In this specific case, `lexinfo:shortForm` could work, as *E* and *ESP* are short forms for *España*. However, this is not always the case, such as *European Assembly* which is the lookup form for *European Parliament*<sup>6</sup>, which could be modelled as a variant; or *Kronkolonie Anguilla* which is the lookup form for *Anguilla*<sup>7</sup>, which could be modelled as a narrower concept.

Still, as observed in Figure 3, additional data may be added to the lookup form, so it might be necessary to propose a class to accommodate this information.

### 3.4 Concept Cross-References

Certain IATE concepts include cross-references; in other words, they provide information regarding the relationships between concepts. In particular, 19 types of relations have been identified, as displayed in Table 1.

Some of the relations have a certain linguistic nature, such as taxonomic ones. These types of

<sup>6</sup><https://iate.europa.eu/entry/result/126540>

<sup>7</sup><https://iate.europa.eu/entry/result/883501>

Cross-reference type	Concept ID
is narrower than	750475
is broader than	901212
is related to	114385
is not to be confused with	1620578
is antonym of	750475
is part of	3588819
has as part	901212
is capital city of	1891420
has as capital city	861168
is currency of	901212
has as currency	861168
is demonym of	1891744
has as demonym	883501
is caused by	1255366
is cause of	3640243
is predecessor of	2246619
is successor of	3591743
is seat of	3630354
has as seat	126540

Table 1: IATE examples of cross-references

relationships can be effectively modelled using SKOS (Simple Knowledge Organisation System)<sup>8</sup>, a structured vocabulary specifically developed for the representation of thesauri. SKOS facilitates the expression of several conceptual relations, including *is narrower than* and *is broader than*. It is important to note that in SKOS terminology, the IATE relation *is narrower than* corresponds to `skos:broader`, whereas *is broader than* aligns with `skos:narrower`. Similarly, the IATE relation *is related to* is represented using the SKOS property `skos:related`. All these SKOS properties require instances of `skos:Concept` in both their domain and range. However, as the class `ontolex:LexicalConcept` is defined as a subclass of `skos:Concept`, no issues of semantic incompatibility arise in this context.

Nonetheless, SKOS alone is insufficient to capture all types of cross-references. For example, it cannot be used to represent antonyms, which are denoted in IATE by the value *is antonym of*. While this relationship cannot be expressed using SKOS, it can be modelled through LexInfo<sup>9</sup>, an ontology that complements the Ontolex framework and provides a set of linguistic data categories. Specifically, the property `lexinfo:antonym` may be employed;

<sup>8</sup><http://www.w3.org/2004/02/skos/core#>

<sup>9</sup><http://www.lexinfo.net/ontology/3.0/lexinfo#>

however, it should be noted that this property is designed to operate between Lexical Senses, rather than at the level of Lexical Concepts. As such, the IATE data structure would need to be adapted this modelling approach. For instance, in IATE, the concepts [750475](#) and [3627400](#) are regarded as antonyms. In order to represent this accurately using `lexinfo:antonym`, the Lexical Sense of *random error* (a term associated with concept [750475](#)) would need to be linked to the Lexical Senses of *systematic error*, *systematic error of measurement*, and *systematic measurement error* (all of which are associated with concept [3627400](#)). Nevertheless, as we would prefer to be the most loyal possible to the original structure, other options will have to be explored.

In addition to linguistic relations, other types of associations may be identified, such as *has as capital city*, *has as currency*, or *has as demonym*, among others. As these relations are extra-linguistic in nature, it is necessary to employ alternative ontologies, such as the DBpedia Ontology (DBO)<sup>10</sup>. For instance, the property `dbo:capital` has been proposed to model the relation *has as capital city*. However, this property imposes constraints on both its domain and range. Consequently, the subject of the triple must be declared as both a `ontolex:LexicalConcept` and a `dbo:PopulatedPlace`. Similarly, the object must simultaneously belong to the classes `ontolex:LexicalConcept` and `dbo:City`.

Similarly, the cross-reference *has as currency* has been proposed to be modelled using the property `dbo:currency`. In order to satisfy the property's constraints, the object of the triple must be classified as a `dbo:Currency`, in addition to being declared a `ontolex:LexicalConcept`.

However, some cross-references, although present in DBO, cannot be utilised due to the nature of the property type. For example, to represent the cross-reference *has as demonym*, the property `dbo:demonym` was identified. Nevertheless, this is a data property, as it takes a string as its object. Nevertheless, the representation of IATE cross-references requires the use of an object property. As a result, the use of Wikidata<sup>11</sup> has been proposed, specifically the property `wdt:P1549`.

To summarise, in order to comply with the constraints imposed by certain properties, it is often

necessary for instances to be assigned an additional class alongside `ontolex:LexicalConcept`. Furthermore, some properties may involve a restructuring of the data model to ensure conformity with their restrictions (e.g., with the property `lexinfo:antonym`). Finally, Table 2 presents a number of proposed solutions for modelling cross-references, although some are still ongoing work and have no modelling suggestions yet.

## 4 Conclusion and Future Work

In this paper, we have explored the application of the Ontolex model to the representation of IATE. While this task has been addressed in previous studies—primarily focusing on parameters such as definitions, notes, sources, and reliability codes—our work has concentrated on modelling other distinctive features of IATE. These include regional usage of terms, the life cycle of entries, lookup forms, and concept cross-references.

On the one hand, we have encountered limitations with certain existing properties, such as in the representation of antonymic relations between concepts. On the other hand, some features of IATE currently appear to be beyond the scope of representation. This is the case, for instance, with the source attribution for regional usage, or the treatment of lookup forms.

As for future work, the representation of several cross-reference types remains an open issue. Likewise, the modelling of the lifecycle of terminological entries is ongoing and may benefit from further discussion. Beyond IATE, we also aim to examine other terminological resources and assess their modelling requirements. One such example is TERMDAT<sup>12</sup>, a database which considers the validation status<sup>13</sup> of a term record.

## Acknowledgments

This work has received funding from the INESData project (Infrastructure to Investigate Data Spaces in Distributed Environments at UPM - TSI-063100-2022-0001, under the UNICO I+D CLOUD call) and TeresIA (Terminology and AI), both funded by the Ministry for Digital Transformation and the Civil Service, within the framework of the recovery plan PRTR financed by the European Union (NextGenerationEU).

<sup>10</sup><https://dbpedia.org/ontology>

<sup>11</sup>[https://www.wikidata.org/wiki/Wikidata:Main\\_Page](https://www.wikidata.org/wiki/Wikidata:Main_Page)

<sup>12</sup><https://www.termdat.bk.admin.ch/>

<sup>13</sup><https://github.com/ontolex/ontolex/blob/master/notes/terminology-requirements.md>

Cross-reference type	Property proposal
is narrower than	skos:broader
is broader than	skos:narrower
is related to	skos:related
is not to be confused with	owl:differentFrom
is antonym of	lexinfo:antonym
is part of	dcterms:isPartOf, rico:isOrWasPartOf, dul:isPartOf
has as part	dct:hasPart, rico:hasOrHadPart, dul:hasPart
is capital city of	
has as capital city	dbo:capital
is currency of	
has as currency	dbo:currency
is demonym of	
has as demonym	wdt:P1549
is caused by	dbo:causedBy
is cause of	
is predecessor of	rico:precedesInTime
is successor of	rico:rico:followsInTime
is seat of	dul:hasLocation
has as seat	dul:isLocationOf

Table 2: Property proposals for IATE cross-reference representation

## References

- Mihael Arcan, Elena Montiel-Ponsoda, John Philip McCrae, and Paul Buitelaar. 2018. Automatic enrichment of terminological resources: the IATE RDF example. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC-2018)*. <https://doi.org/10.5281/zenodo.2599942>.
- Julia Bosque-Gil, Jorge Gracia, Elena Montiel-Ponsoda, and Guadalupe Aguado-de Cea. 2016a. Modelling multilingual lexicographic resources for the web of data: The k dictionaries case. In *GLOBALEX 2016 Lexicographic Resources for Human Language Technology Workshop Programme*, page 65.
- Julia Bosque-Gil, Elena Montiel-Ponsoda, Jorge Gracia, and Guadalupe Aguado-de Cea. 2016b. Terminoteca RDF: a Gathering Point for Multilingual Terminologies in Spain. *Term Bases and Linguistic Linked Open Data*.
- Caterina Caracciolo, Armando Stellato, Ahsan Morshed, Gudrun Johannsen, Sachit Rajbhandari, Yves Jaques, and Johannes Keizer. 2013. The agrovoc linked dataset. *Semantic Web*, 4(3):341–348.
- Philipp Cimiano, John P McCrae, Víctor Rodríguez-Doncel, Tatiana Gornostay, Asunción Gómez-Pérez, Benjamin Siemoneit, and Andis Lagzdins. 2015. Linked terminologies: applying linked data principles to terminological resources. In *Proceedings of the eLex 2015 Conference*, pages 504–517. ISBN 978-961-93594-3-3.
- Maria Pia di Buono, Philipp Cimiano, Mohammad Fazleh Elahi, and Frank Grimm. 2020. *Terme-à-LLOD: Simplifying the conversion and hosting of terminological resources as linked data*. In *Proceedings of the 7th Workshop on Linked Data in Linguistics (LDL-2020)*, pages 28–35, Marseille, France. European Language Resources Association.
- Luisa Alvite Díez, Beatriz Pérez-León, Mercedes Martínez-González, and Dámaso-Javier Vicente Blanco. 2010. Propuesta de representación del tesoro Eurovoc en SKOS para su integración en sistemas de información jurídica. *Scire: representación y organización del conocimiento*.
- Jorge Gracia, Marta Villegas, Asuncion Gomez-Perez, and Nuria Bel. 2018. The Apertium bilingual dictionaries on the web of data. *Semantic Web*, 9(2):231–240.
- Patricia Martín-Chozas, Thierry Declerck, Elena Montiel-Ponsoda, and Víctor Rodríguez-Doncel. 2024. *Representing terminological data in the semantic web*. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication*.
- John P McCrae, Julia Bosque-Gil, Jorge Gracia, Paul Buitelaar, and Philipp Cimiano. 2017. The Ontolex-Lemon model: development and applications. In *Proceedings of eLex 2017 conference*, pages 19–21.
- John P. McCrae, Dennis Spohr, and Philipp Cimiano. 2012. *The lemon Lexicon Model for Ontologies*. In *Proceedings of the 3rd International Conference*

*on Semantic Web Applications and Tools for Life Sciences (SWAT4LS).*

Joachim Neubert. 2009. Bringing the "Thesaurus for Economics" on to the Web of Linked Data. *LDOW*.

Juan-Antonio Pastor-Sánchez. 2016. Proposal to represent the unesco thesaurus for the semantic web applying iso-25964. *Brazilian Journal of Information Studies: Research Trends*, 10(1):1–8.

Benjamin Zapilko, Johann Schaible, Philipp Mayr, and Brigitte Mathiak. 2013. TheSoz: A SKOS representation of the thesaurus for the social sciences. *Semantic Web*, 4(3):257–263.

Paula Zorrilla-Agut and Thierry Fontenelle. 2019. IATE 2: Modernising the EU's IATE terminological database to respond to the challenges of today's translation world and beyond. *Terminology*, 25(2):146–174.